Ch. 22 notes

2 proportion Int/test

Example: Do men and women wear seatbelts the same % of the time? In order to find out, we take an SRS of 200 male drivers and another SRS of 250 female drivers. We find that there are 146 men that wear their seatbelts regularly and 203 women that wear them regularly. Is there a significant difference between the two genders? What is the difference (if any)? (a, b)

$H_0: p_M = p_F$

$H_a: p_m > < \neq p_F$

$\hat{p}_M = \frac{146}{200}$ $\hat{p}_F = \frac{203}{250}$

$\hat{p}_m = 0.73$ $\hat{p}_F = 0.812$

Ch. 22: 2 Proportion Z Interval and Test

Conditions (for both interval and test)

(Double the conditions from one proportion)

1) 2 independent SRS

1) Male & female drivers are indep. of each other. SRS of both stated

2) $n_1\hat{p}_1 \geq 10$ $\quad$ $n_2\hat{p}_2 \geq 10$

   $n_1\hat{q}_1 \geq 10$ $\quad$ $n_2\hat{q}_2 \geq 10$

3) $pop_1 \geq 10n_1$

   $pop_2 \geq 10n_2$

$N(\, , \,)$

Conditions met -> Normal Model -> 2 prop Z Int (or Test)

2-Prop Z-Interval:

* Since the normal model is being used, we need a mean and a std. deviation (std. error)

$$N\left(\hat{p}_1 - \hat{p}_2, \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right)$$

* Mean: $\hat{p}_1 - \hat{p}_2$

* Std. Error: (think, combining two variables, X - Y)

$$\sigma_{M-F} = \sqrt{\sigma_m^2 + \sigma_F^2} = \sqrt{\frac{\hat{p}_M\hat{q}_M}{n_M} + \frac{\hat{p}_F\hat{q}_F}{n_F}}$$

* Confidence Interval Formula: statistic ± *(crit. value)(std. error)*

diff.

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = (a, b)$$

2prop Z Int

III. Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

Confidence interval: statistic $\pm$ (critical value) $\cdot$ (standard deviation of statistic)

Single-Sample

Statistic	Standard Deviation of Statistic
Sample Mean	$\dfrac{\sigma}{\sqrt{n}}$
Sample Proportion	$\sqrt{\dfrac{p(1-p)}{n}}$

Two-Sample

Statistic	Standard Deviation of Statistic
Difference of sample means	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$
Difference of sample proportions	$\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ Special case when $p_1 = p_2$ $\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

Conclusion:

We are __95__% confident that the difference in the proportions of __male____ and _female ~~seatbelts~~ wearers_ is between __5%____ and __10____%.

$(0.05, 0.10)$

*How would we know if the 2 proportions are the same? What would the difference be?

$H_0: p_M = p_F$

$(-0.02, 0.08)$

$0 = p_1 - p_2$

Example:

College students were randomly selected and asked about how much alcohol they consumed on a weekly basis. Over a certain amount of alcohol consumed was considered binge drinking. Out of 5348 males surveyed, 1392 were identified as binge drinkers. Out of 8471 females surveyed, 1748 were identified as binge drinkers. What is the difference in the proportions of male and female binge drinkers? Use 95% confidence. ~~Assume conditions met.~~

$\hat{p}_m = 1392/5348 = 0.2603$

$\hat{p}_F = 1748/8471 = 0.2064$

Conditions:

1) 2 indep SRS

1) stated random samples
Male & female college students are indep. of each other

2) $n_M \hat{p}_M$, $n_M \hat{q}_M$, $n_F \hat{p}_F$, $n_F \hat{q}_F \geq 10$

2) 1392, 3956, 1748, 6723 ≥ 10

3) $pop_m \geq 10 n_m$
$pop_F \geq 10 n_F$

3) There are more than 53,480 male college students & 84,710 female college students

Conditions met → Normal model → 2 prop Z Interval

$$\underset{M\ -\ F}{(0.2603-0.2064)} \pm (1.96)\sqrt{\frac{(0.2603)(0.7397)}{5348}+\frac{(0.2064)(0.7936)}{8471}}$$

$$=(0.03935, 0.06851)$$

We are 95% confident that the true difference in the percent of male & female binge drinkers is btw. 3.935% and 6.851%.

2-Proportion Z-test:

We can compare 2 proportions together

Hypotheses:

Ho: $p_1 = p_2$ $\qquad$ $H_0: p_1 - p_2 = 0$

Ha: $p_1 >, <, \neq p_2$ $\qquad$ $H_a: p_1 - p_2 >, <, \neq 0$

Conditions:

* Same as confidence interval $\qquad$ Calc.
* Conditions met -> Normal model -> 2-Prop Z-Test

Mechanics:

- Recall: Ho: $p_1 = p_2$
- We must assume Ho is true until proven otherwise
- So if $p_1 = p_2$, then we should really combine $\hat{p}_1$ and $\hat{p}_2$ so that we have more info about true p.
- This is called ***pooling*** our sample proportions together

$\hat{p}_1 = \dfrac{X_1}{n_1}$ $\qquad\qquad$ $\hat{p}_2 = \dfrac{X_2}{n_2}$ $\qquad\qquad$ $\hat{p}_1 = \dfrac{70}{100} = 0.70$ $\qquad$ $\hat{p}_2 = \dfrac{160}{200} = .80$

$$\text{Pooled } \hat{p} = \frac{X_1 + X_2}{n_1 + n_2} \neq \hat{p}_1 + \hat{p}_2$$

$$\hat{p} = \frac{70+160}{100+200}$$

Mean: $\hat{p}_1 - \hat{p}_2$

$$N\left(\hat{p}_1 - \hat{p}_2, \qquad \right)$$

Std. Error (pooled):

$$\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$$

$H_0: p_1 = p_2$

$p_1 - p_2 = 0$

Test Statistic:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

III. Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

Confidence interval: statistic $\pm$ (critical value) $\cdot$ (standard deviation of statistic)

Single-Sample

Statistic	Standard Deviation of Statistic
Sample Mean	$\dfrac{\sigma}{\sqrt{n}}$
Sample Proportion	$\sqrt{\dfrac{p(1-p)}{n}}$

$\underline{\hat{p}_1 - \hat{p}_2}$

Two-Sample

Statistic	Standard Deviation of Statistic
Difference of sample means	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$
Difference of sample proportions	$\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ Special case when $p_1 = p_2$ $\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

P-value: same as always   P(Z $\gtrless$ test statistic)

Conclusion:

- Same first sentence:
  We reject/fail to reject b/c p-value...

- Second sentence:
  We have sufficient/insufficient evidence that the percent of __(#1)___ is __$>$ $<$ $\neq$__ to the percent of __(#2)___.
  (Ha)

Example: Back to the example from the beginning: Do men and women wear seatbelts the same % of the time? In order to find out, we take an SRS of 200 male drivers and another SRS of 250 female drivers. We find that there are 146 men that wear their seatbelts regularly and 203 women that wear them regularly. Is there a significant difference between the two genders? (assume conditions met)

$\hat{p}_m = 146/200 = 0.73$

$\hat{p}_F = 203/250 = 0.812$

Cond. met → Normal model

2 prop z test

$H_0: p_m = p_F$

$H_a: p_m \neq p_F$

$$Z = \frac{0.73 - 0.812}{\sqrt{\frac{(0.776)(0.224)}{200} + \frac{(0.776)(0.224)}{250}}}$$

$$\hat{p} = \frac{146 + 203}{200 + 250}$$

$$\hat{p} = 0.776$$

$$Z = -2.0717$$

$$2 \cdot P(Z < -2.0717) = 0.03829$$

$\alpha = 0.05$

We reject H_0 b/c p-value of $0.03829 < \alpha = 0.05$. We have sufficient evidence that the percent of males who wear seatbelts is not equal to the percent of females who wear seatbelts.

p. ~~519~~ 520 #9, 17

Do conditions on #9 only

Cond. met on #17

9) (a)

STATE	CHECK
- 2 indep. SRS	- stated random samples and senior men & women are indep. of each other
- $n_M \hat{p}_M$	- 411
$n_M \hat{q}_M \geq 10$	- 601 ≥ 10
$n_W \hat{p}_W$	- 535
$n_W \hat{q}_W$	- 527
- $pop_M \geq 10 * n_M$	- there are more than 10120 senior men
$pop_W \geq 10 * n_W$	and more than 10620 senior women

conditions met -> Normal Model -> 2 prop Z Interval

(b) $\hat{p}_m = 411/1012$ $\hat{p}_w = 535/1062$

$$(0.406 - 0.504) \pm 1.96\sqrt{\frac{(0.406)(0.594)}{1012} + \frac{(0.504)(0.496)}{1062}}$$

$= (-0.1403, -0.055)$

We are 95% confident that the true difference in the % of senior men and women who have arthritis is between 14.03% and 5.5%.

(c) Yes, our interval suggests a difference between the % of males and females who suffer from arthritis. The value of 0 (no difference between the 2 proportions) is NOT in the interval, suggesting that there is a difference.

Since our interval is entirely negative, and we subtracted MEN - WOMEN, this indicates that WOMEN have the higher % of arthritis sufferers.

17) $\hat{p}_D = 54/284$ $\hat{p}_L = 11/41$

(a) Prospective Study with Blocking

(b) Ho: $p_D = p_L$

Ha: $p_D \neq p_L$

(b) <u>STATE</u>	<u>CHECK</u>
- 2 indep. SRS	- assumed representative samples and parents are indep. of each other
- $n_D\hat{p}_D$	- 54
$n_D\hat{q}_D \geq 10$	- 230 ≥ 10
$n_L\hat{p}_L$	- 11
$n_L\hat{q}_L$	- 30
- $pop_D \geq 10*n_D$	- there are more than 2840 kids whose parents disapprove and more than 410 whose parents are lenient on smoking
$pop_L \geq 10*n_L$	

conditions met --> normal model for 2 prop Z test

(d) $$Z = \frac{0.1901 - 0.2683}{\sqrt{\frac{(0.2)(0.8)}{284} + \frac{(0.2)(0.8)}{41}}} = -1.169$$

2*P(Z < -1.169) = 0.2422

We fail to reject Ho b/c p-value of 0.2422 is > α = 0.05.
We have insufficient evidence that the percent of students who started smoking is different between lenient and disapproving parents.

(d) There is a 24.22% of getting a sample where the diff. btw. the % of smokers in lenient & disapproving households is 7.8% or larger, if the %s are really equal.

$\hat{p}_D - \hat{p}_L$

(e) Concl: fail to reject

Type II